DOCUMENT RESUME

ED 329 302 JC 910 120

AUTHOR Plucker, Frank E.
TITLE The Use and Validation of Qualitative Methods Used in Program Evaluation.
PUB DATE Apr 90
NOTE 16p.; Paper presented at the Annual Research Conference of the California Association of Community Colleges (Asilomar, Monterey, CA, April 18-20, 1990).
PUB TYPE Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Community Colleges; Data Interpretation; *Educational Research; English Curriculum; Generalization; Program Evaluation; *Qualitative Research; *Research Methodology; Two Year Colleges; *Validity

ABSTRACT

When conducting a two-year college program review, there are several advantages to supplementing the standard quantitative research approach with qualitative measures. Qualitative research does not depend on a large number of random samples, it uses a flexible design which can be refined as the research is executed, and it generates findings in a manner more accessible to policymakers and practitioners. Previous research comparing qualitative validity measures with quantitative validity suggests that new categories for validity research are needed. Specifically, the terms "local validity" and "general validity" have been proposed for assessing the validity of research results. Local validity addresses the validity for a particular set of conditions of a study, while general validity refers to the extension of the study results beyond the particular set of conditions. A review of the English program at College of Alameda (California) utilized both qualitative and quantitative measures. Guided discussions, interviews, and informal conversations between the evaluator and students, counselors and teachers were recorded in field notes or "contact summaries." In addition, survey instruments were administered to faculty, students, and the program administrator. To validate study results, three methods were used: triangulation (finding at least three sources to corroborate a viewpoint); saturation (developing new understandings from the inquiry process); and member checks (having study participants review conclusions). Program review conclusions generated by the qualitative method, tables illustrating the methods for establishing validity used by educational researchers, and the contact summary forms are included. (GFW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED329302

THE USE AND VALIDATION OF QUALITATIVE METHODS USED IN PROGRAM EVALUATION

FRANK E. PLUCKER, ASSISTANT DEAN
DIVISION OF ARTS AND LETTERS
COLLEGE OF ALAMEDA

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

F. E. Plucker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

JC 910120

Presented at:

CACC Annual Research Conference
Asiloman

April 18, 19, & 20, 1990

BEST COPY AVAILABLE

The Use and Validation of Qualitative Methods used in Program Evaluations

Division of Arts and Letters
College of Alameda

Introduction:

The purpose of this English program review was to determine the strengths and weaknesses of the program and what was needed to correct these weaknesses or improve the program. With an understanding of the program strengths and weaknesses based on valid research processes the administrator can promote what is effective and of high quality and fight for funds to strengthen the weak areas. A good program review dissects the structure into the functional areas of support, such as administrative support, operational funding, faculty effectiveness, curriculum and student placement. The College of Alameda (COA) utilized a review process incorporating both qualitative and quantitative measures. Qualitative measures were used to provide information not readily available from quantitative measures such as quality and effectiveness of the program.

There are several advantages to using qualitative methods in program review. First it is assumed that the observations and data will be collected in context (classroom, organization) not a laboratory and views the situation holistically rather than in components. Second, while the human limitations are accepted, advantages are used to improve the observations through tacit knowledge and flexibility to pursue facts and concepts as they arise. Third, qualitative research accepts the fact that all research is subjective and human (Bronowski, 1971; Lincoln, 1985). Fourth, it does not depend on large number random samples, an advantage, since true random samples are virtually impossible in the real world. Fifth, qualitative research is not limited by the design, but uses an emerging design that is dynamic and developed and is refined as the research is executed. Finally, quoting Miles and Huberman (p 15, 1984)

> the findings from qualitative studies have a quality of "undeniability". Words, especially when they are organized into incidents or stories, have a concrete, vivid, meaningful flavor that often proves far more convincing to a reader, another researcher, a policy maker, a practitioner, than a page of numbers.

Principles of Validating Qualitative Technique

Plucker and Sabbers (1987) made a comparison of qualitative validity measures with quantitative validity measures. This comparison was made to clarify the type of validity obtained with a particular qualitative method. It also provided a reference point for those familiar with the accepted terminology of experimental design.

It was their view that qualitative and quantitative are two approaches to investigation by humans who are neither purely rationalistic or naturalistic. The types of evidence presented by qualitative and quantitative methodologists might be much more compatible than a reading of the literature would lead one to believe. Many of the differences that are stated could be due to different vocabulary rather than different beliefs. True, there is a great difference in the approach that one takes to an investigation if Verstehen rather than Positivism (Monette, Sullivan, and DeJong, 1986) is viewed as the proper orientation to study behavior. They did not intend to disagree with Smith and Heshusius (1986) who contended that the differences between the approaches should not be resolved; rather each approaches should exist independent of the other. They did not believe that any investigation in educational research would be "value free" which is a requirement under the rationalistic paradigm; nor did they believe that all reality is as divergent as is often argued by those with a naturalistic viewpoint.

The analysis reported in their presentation was based on their position that when an investigator attempts to describe a belief or finding, there is a certain logic needed to convince the reader of the reasonableness of the inference. One would expect that many convincing arguments would follow similar paths of logic. There is not a priori reason for the logic to be different for the qualitative than for the quantitative methodologist, or for the researcher employing large sample sizes than for the one studying single organisms. It is for the reader that the finding is based on "research evidence" rather than casual personal experience. Because there is reason to expect similarities in the logic needed for the many different approaches to research, it is to be expected that there is a scheme within which these similarities can be compared.

Plucker and Sabbers made the following assumptions:

1) There is some transferability of understanding research studies in education from one situation to another. This is a necessary requirement to build a body of knowledge.

2) There are methods to determine the degree of validity of a statement (conclusion or contention) resulting from any type of quantitative or qualitative study. As Ziman stated, "The goal of science is a consensus of rational opinion over the widest possible field" (1978, p. 3).

3) The real task in evaluating a research project is in determining the degree of validity of a conclusion from the data presented in the study. The main goal of a published research

project is the same as the main goal of Borg's (1987) book, that is, "to make the reader an intelligent consumer of educational research" (p 1).

They examined the similarities among methods of validating research findings as discussed in the following sources: Campbell and Stanley (1963), Cook and Campbell, (1979), Miles and Huberman (1984), Lincoln and Guba (1985), and Smith and Heshusius (1986). In addition to these sources,, discussions with researchers regarding their beliefs and their understanding of the methodologies proposed by the above authors were conducted. There was a concerted effort to represent the viewpoints of both orientations to data analysis in this work; however, we agree with Guba (1987) that there should be no attempt to find an equivalence between the naturalistic and conventional paradigms.

New Categories for Research Validity

Plucker and Sabbers (1987) proposed new categories for the following reasons. First, to eliminate the meanings of internal and external validity used by Campbell and Stanley (1963) because these terms have been refined by Cook and Campbell (1979). Second, to replace the words used by Cook and Campbell because those in the field may have a preconceived concept of the standard categories of internal validity, statistical conclusion validity, construct validity, and external validity. Instead, they suggest these terms are special cases useful for quasi - and experimental research. Finally, we believe the proposed terms better describe the meaning to be conveyed by the validity types. (Comment-Hindsight approaches 20-20; e.g., if Cook and Campbell had read Borg's (1984) article prior to 1979 they would have had a different presentation of the threats that randomization does not rule out. Thus, we intend no criticism of Campbell and Stanley for what they did not know in 1963).

The proposed categories to replace Campbell and Stanley's two terms are local validity and general validity. Local validity would address the validity for the particular set of conditions of the study. General validity would refer to the extension of the study results beyond the particular set of conditions.

It is further proposed that two categories under local validity be conclusion validity and internal validity. Under general validity, two categories, construct validity and external validity should be included. Definitions of these terms are proposed as follows.

1 Conclusion Validity

Conclusion validity does not need the restrictive name of "statistical conclusion validity". "Statistical" implies large

BEST COPY AVAILABLE

sample research, or at least rules out the typical qualitative data analysis or single subject research. The general category of conclusion validity includes any attempt to remove nonsystematic error from a finding to indicate that what is being interpreted is not a chance occurrence.

Cook and Campbell (1979) state that conclusions about covariation are made on the basis of statistics. This condition precludes most qualitative research from ever attaining statistical conclusion validity. However, covariation is a prerequisite for causal statements, but there is a great amount of important research, especially in education, that is not intended to show causation. Also, studies concerned with causation need not have quantitative data subject to tests of covariation. Therefore, it is easy to visualize a broader category concerned with methods of ruling out chance. Ziman's (1978) statement that science is built on "intersubjectivity" supports the contention that the data within a study should lead to a local consensus. Thus one description for conclusion validity would be a consensus within the data.

For single subject research, conclusion validity asks, "was the baseline stable prior to treatment intervention" and "is there a true reversal across interruption, or a true change from one phase to another?" For qualitative studies, the question may be, "Was the researcher there long enough to really know what is happening or was this a one-time occasion that is being talked about?"

Conclusion validity includes any data to rule out the chance happening leading one to interpret something other than a true relationship between variables or a systematic occurrence. Therefore it is clear that the requirement that "covariation exists between variables" from Cook and Campbell is a special case of conclusion validity.

2. Internal Validity

Internal validity addresses whether the interpretation being made is correct for the case at hand. This category includes all attempts to remove systematic error (bias) from an interpretation. There is no need to include causation in this category, as doing so restricts our use to experimental and quasi-experimental research. We ask whether the statement adequately explains what happened in this instance, given the subjects studied and the variables as measured. In Wolcott's conclusion that children were the best data gatherers in the community, we would restrict the conclusion to the data reported in their reports for the community in which he taught.

Previous definitions of internal validity (cf. Cook and Campbell, 1979) examines causal relationships. The methods of

identifying causality for the methods of Miles and Huberman and Guba & Lincoln are included in the tables. Many of these same methods are used for descriptive findings where no causality is inferred. For example, a member check to establish the interpretative accuracy of an inference is relevant in a descriptive study. Likewise, eliminating rival hypotheses is practiced by both qualitative and quantitative methodologists.

A description of internal validity would be a consensus of interpretation within the study.

3. Construct Validity

With construct validity we eliminate the "putative causes and effects" and restrict our reference to the variables to which we might generalize the results. The construct validity category is asking about the "construct validity" as discussed in the educational measurement literature. The category is concerned with the extent to which we infer the psychological/social constructs measured by the variables in the study. Construct validity is established when the consensus is that the best of all possible psychological/social constructs fits the data far better than the second best.

In Wolcott's example, one might ask if children would be the best sources of information about the economic trends in the community. The generalization would be to a different construct within the same setting. For construct validity, one would ask whether there was an adequate operationalization of the construct to which one wishes to generalize.

4. External Validity

External validity answers the question "How do our conclusions generalize to an across other populations, situations, and times". There is really no difference introduced when we eliminate the causation emphasis of Cook and Campbell. It is true that generalization is often not considered to be an important part of interpretative research. The reader can determine the appropriate generalization if the writer presents a clear description of the population, setting, and occasion studied. This criterion for generalization works equally well for experimental laboratory work or descriptive research.

To establish external validity of an inference made from a study, one must examine the similarities and differences of populations, settings, and times which might affect the "truth" of the statement.

The data obtained from an analysis of the content contained

in the sources listed above are presented in Tables 1 and 2. The actual data presented in these tables are the descriptors of categories of information. These descriptors are used in the sources cited above by the authors in their explanation of their methods to establish research validity.

The similarities in the explanations of the methods presented by the authors of the cited sources lead Plucker and Sabbers to conclude that the logic used for the two methodologies is very similar. However, the categories of research evidence is very arbitrary. For example, how triangulation is used in a particular instance could warrant a placement into conclusion validity whereas another use may be classified as internal or construct.

The current draft of the tables is provided to clarify for the reader methods of establishing validity used by educational researchers. Plucker and Sabbers were convinced of a great similarity in the evidence needed to substantiate claims of validity by both qualitative and quantitative researchers.

Discussion of the Validity Categories

Note that just like anything else concluded about validity, these categories are restricted to statements rather than to studies. Other conclusions of Wolcott might be treated completely different from the one used as an example here.

Much of the apparent difference in research validation stem from the specificity of the area within which the authors present their "models". For example, Cook and Campbell's title delimits "Quasi-experimentation" as the realm within which their categories are discussed. Campbell and Stanley restricted their discussion to "Experimental and quasi-experimental" designs. Treatments of internal validity, especially rival hypotheses or threats to internal validity as presented by Campbell and Stanley, are limited in most textbooks to chapters on experimental (and/or quasi-experimental) designs. The student reading such a text has no conception that rival hypotheses are as big a threat to conclusions drawn from single-subject studies or from survey or correlational research, and need to be considered whether qualitative or quantitative methods are used.

The topics addressed in the presentation of Plucker and Sabbers is presented here to introduce college institutional researchers to the similarities of evidence needed in many research endeavors. It is also intended to illustrate how "qualitative" methods may be used and validated in an institutional setting.

Methodology for this Program Review

Data for this program review was gathered from a variety of

sources. The evaluator was immersed in the program through faculty discussions, faculty meetings, individual contacts and directed discussions. In addition purposeful contacts were made with students and counselors. Data from these types of contacts was recorded and kept using diaries, field notes or contact summaries. All records need specific information on the date, type and purpose of the contact along with quotes and conversation supplying information needed in the research. I have found contact summaries invaluable for these records. A form should be prepared for the contact summaries (see examples in the appendix). Validation of qualitative data requires good documented observations. The key to data collection in all scientific investigations is documentation.

Data was also gathered through the use of surveys administered to faculty, students and the program administrator. The student survey consisted of a series of questions with multiple choice answers or Likert scales. At the end additional comments were requested. The faculty survey consisted of several scaled questions on student preparation, adequate scheduling, administrative support and institutional support. This was followed with a series of open ended questions concerning the English program. In addition the faculty were requested to complete a unit plan for the English program. Unit plans are requested in all the programs and provide the main source for developing the divisional three year plan. The unit plan asks for an assessment of the current status, identification of the major problems, what needs to be done to overcome the problems and a projection of where the program should be three years from now.

Many discussions and individual conversations occurred during the process of the program evaluation. Casual conversations can provide a data source when documented. One method is to gather data in interviews or in conversations by using comments (Snow, Zurcher and Sjoberg, 1982). Interviewing by comment involves the inquirer to make a comment that elicits a response. The same method is used in conversations. The comment can be either in agreement with the current understanding or in conflict. Comments indicating that the interviewer has an incorrect understanding usually illicite responses correcting the inquirer. For example, during the English program review process one problem identified was the placement of students in the appropriate level class. A comment could be "Now that we are doing XXX in the matriculation process students are being accurately placed." Comments are effective when integrated into interview schedules (Measor, 1985). Interview schedules and guided discussions combined with contact summaries provide an efficient way of collecting interview and discussion data.

Validation Techniques

Three methods of validation were used in this program review,

triangulation, saturation and member checks. Triangulation involves finding three or more methods or sources with the same understanding of or viewpoint. Saturation is the point at which now new understanding is developed from the inquiry process. Member checks involve the participants in the study by having them review the conclusions reached by the inquirer.

Faculty surveys, student surveys, conversations with the counsel and faculty and the administrators evaluation provided data source for triangulation. One conclusion of this study was that COA needed a more effective ESL program. The was verified by faculty comments in conversations and on the faculty survey, conversations by the counselors, student comments on the student survey and the administrators evaluation. Thus there were four sources with the same understanding.

Saturation can be illustrated using the conclusion from the study "limited class size of 30 students". As might be expected the faculty were singularly together in their opinion on class size. The administrator and the students were also in agreement with this opinion for composition classes. This is in agreement with reason: For example, a class with 30 students has a faculty member with 3 sections, and the students produce one paper per week for grading. Using a conservative 15 minutes per paper requires 22 1/2 hours per week. If the instructor provides comments and other needed input to the student more than 15 minutes per paper will be required.

The final validity check had the faculty review the entire program evaluation and make comments or point out error in understanding. This is one method of making a member check. To check to see if the members of the group in the research project have the same understanding of the situation or problem. The English faculty were generous with their contribution to the wording, phrasing and grammatical correction. They did however concur with the conclusions.

Advantages of Qualitative methods

The conclusions of the English program review were:

Strengths

a) College level English program, the faculty, students, counselors and the division administrator agreed that the college level courses were appropriate and effective.
b) The faculty were endorsed as effective by the students, administrator, counselors and the faculty themselves.

Weaknesses

a) Students were enrolled into college level and preparation

BEST COPY AVAILABLE

courses with insufficient skill to succeed. This was the opinion of faculty, counselors and the administrator.
b) The precollege level English and ESL courses were not adequately designed and articulated. This is based on comments by students, faculty and counselors.

Areas for improvement

a) The department needs to develope an English as a second language program and hire full time faculty to support it.
b) The class size of composition classes should be limited to 30 and more sections scheduled to meet the needs of the students.
c) There was a need for an organized series of courses preparing students for college level writing courses and this should be coordinated with counseling and placement.

Pure quantitative data by itself would not support these conclusions. This is particularly true of the class size conclusion. Most productivity measure are in direct conflict with this conclusion. The need for more contract faculty is expressed in AB1725, not any measured need to improve the program. This study concludes that more contract faculty are needed to improve the program quality not just offer more sections. We are all continually using qualitative inputs to arrive at recommendations, develop inferences or define questions that are proven with quantitative projects. One can find some quantitative measures that would indirectly infer that students were not learning effectively in large classes.

What qualitative validation does is provide experimentally sound methods of validating our research conclusions.

References

Borg, W. R. (1987). Applying educational research (2nd ed.). New York: Longman.

Borg, W. R. (1984). Dealing with threats to internal validity that randomization does not rule out. Educational Researcher, 13, 11-14.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.

Guba, E. G. (1987). What have we learned about Naturalistic Evaluation? Evaluation Practice, 8, 23-43.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.

Measor, Linda, "Interviewing: A Strategy in Qualitative Research," in Strategies of Educational Research: Qualitative Methods , Edited by Robert Burgess, Lewes: The Falmer Press, 1985.

Miles, M. B., & Huberman, A. M. (1984). Qualitative data analysis: A sourcebook of new methods. Beverly Hills, CA: Sage.

Monette, D. R., Sullivan, T. J., DeJong, D. R. (1986). Applied Social research. New York: Holt, Rinehart and Winston.

Plucker, F. E. and Sabbers, D. "A Common Validation Logic for Qualitative and Quantitative Research" AERA National Meeting, Washington DC, April, 1987.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation; The end of the quantitative-qualitative debate among educational inquirers. Educational Researcher, 15, 4-12.

Snow, David A., Zurcher, Louis A., and Gideon Sjoberg, "Interviewing by Comment: An Adjunct to the Direct Question," Qualitative Sociology, 5, 4:285-311, 1982.

Ziman, J. (1978). Reliable knowledge. Cambridge: Cambridge University Press.

Appendices

1) Table 1 Comparison of Validity Terms
2) Table 2 Comparison of Processes for Establishing Validity
3) Contact Summary Form - English Program Review
4) Contact Summary Form - Faculty Motivation

****

Table 1

Comparison of Validity Terms

	Cook/Campbell	Lincoln/Guba	Miles/Huberman
	Statistical conclusion		
LOCAL VALIDITY		Credibility	
	Internal		
			Verifying conclusions
	Construct		
		Transferability	
GENERAL VALIDITY			
	External		

Table 2

Comparison of Processes for Establishing Validity

To Establish\	Cook/Campbell	Lincoln/Guba	Miles/Huberman
Conclusion Validity	Tests of Significance Power analysis	Prolonged engagement Presistent observation *Triangulation Negative case analysis	Checking for represen- Weighting the evidence Making contrasts/comparisons Looking for negative evidence
Conclusion /Internal		Referential adequacy	Checking the meaning of outliers using extreme cases
Internal Validity	Analyze rival hypotheses		Checking for researcher effects Checking out rival explanations
Internal /Construct		*Triangulation Peer debriefing Member check	Triangulation Ruling out spurious relations
Construct Validity	Analyze threats to construct validity		Getting feedback from informants
External Validity	Identify settings, persons, types, times	Develop thick description Do theoretical/ purposive sampling	Replicating a finding

Contact Summary Form

Contact type Project-English program evaluation

1) What were the major issues or themes?

2) Summarize the information on each of the major topics.

a) Class size

b) Student preparation

c)

3) What struck you as salient, interesting on insightful?

4) What new question does this contact raise for use in the next contact?

5) Illustrative quote, statement or example.

Contact Summary Sheet - Motivation Faculty

Contact date_____
Site - I R

What were the main issues or themes?

Where did the interviewer exhibit the most enthusiasm, excitement, or interest?

Who developed (source of) the research agenda?

Grants and Contracts office function?

Relationship between research activity and departmental status?

What is the influence of peers outside the university on individual status?

What was the Causality orientation, Autonomy, Control, Impersonal, or research, funding?

Illuminating or salient obseravations.

Memorable Quotes--

ERIC Clearinghouse for Junior Colleges

APR 05 1991